A SCALE INVARIANT DISTANCE MEASURE FOR TEXTURE RETRIEVAL

R.J. O'Callaghan and D.R. Bull

Image Communications Group,
Centre for Communications Research, University of Bristol,
Woodland Road, Bristol, BS8 1UB, U.K.

ABSTRACT

We propose a similarity measure between two textures based on moments of the Fourier magnitude spectrum. The resulting distance is robust to changes in scale as well as to spatial shifts and grey-scale transforms of the texture samples. This type of invariant distance has applications to content-based image retrieval and classification tasks. We test the performance of the algorithm in a retrieval scenario using texture patches from the Brodatz album. The results indicate that the distance measure emulates human similarity perception in comparing textures.

1. INTRODUCTION

The problems of texture analysis and classification are well-studied ones in the fields of computer vision and image processing. This is not surprising, as there is a wealth of evidence to show that the human visual system is able to exploit textural information for a variety of tasks, from pattern recognition to motion perception. As a result, representations of texture have become, along with colour statistics, almost ubiquitous content descriptors in image and video retrieval systems.

Of course, analysis of natural textures is not a straightforward task. Even if we restrict our set of textures to two-dimensional intensity patterns (i.e. grey-scale images of textured flat surfaces, viewed from a constant angle) we must still contend with variations caused by rotation, scaling and grey-scale transform. Much of the work in the literature applies to rotationally invariant texture analysis, but considerably less to scale invariance. This is surprising, since all naturally imaged textures are subject to scale transformation, due to distance from the camera (or eye), as well as optical (or digital) zooming. On the other hand, many such textures are either intrinsically isotropic or are predominantly viewed in a characteristic orientation (e.g. brick texture). Additionally, recent experimental results have shown that the human visual system can utilise scale information in the perception of visual expansion, without estimation of optic flow [1]. This seems to imply that we have both mechanisms that are sensitive and insensitive to texture scaling (since obviously we recognise similar textures at varying distances) and use both in important perceptual functions. This provides further motivation for the development of computer vision algorithms dealing with texture scale.

Of the existing work dealing with scale invariance, various theoretical approaches have included Fourier-Mellin type transforms [2], wavelet based features [3], random fields [4] and the use of fractal dimension [5]. In this work, we adopt a Fourier transform based approach. It is widely accepted that the human visual system performs some Fourier-type analysis on optical input. It has also been observed that we are unable to discriminate between textures that agree in their second-order statistics [6]. For two-tone textures, this means we cannot distinguish between patterns having identical power spectra. Two computational conveniences are also afforded by Fourier magnitude spectra: there is a relatively simple relationship under scaling and there is no need for resampling, as is the case with log-polar spectra. Specifically given a Fourier transform pair, $f(x,y) \leftrightarrow F(v,\omega)$ we have:

$$f(\alpha x, \alpha y) \leftrightarrow \frac{1}{\alpha^2} F(v/\alpha, \omega/\alpha) \qquad (1.1)$$

2. METHOD

A moment-based approach was used to describe the magnitude spectra, inspired by previous work by Taubin and Cooper [7] on geometric invariants for shape recognition. This theory has already been successfully applied for illumination-invariant colour object recognition, by us and others [8], [9]. A similar, moment based approach was used by Yoshida and Wu [10] for rotation-scaling invariant texture classification. This work differs from theirs, in that rather than develop scale-invariant descriptors of a single texture, we define a scale-

invariant distance measure between a pair of textures. In doing so, we attempt to discard as little information as possible about the frequency distributions. The problem is posed as a texture matching/retrieval task, rather than as a classification task, although the invariant distance could be used for minimum-distance type classification, given a suitable set of exemplars of each class.

Firstly we will introduce, without discussion, the Taubin and Cooper moment matrices [7]. In the theory, centred moments were developed. However, since we do not require invariance to translation of the spectrum (only scaling), our moments are not centred. The two matrices used in the algorithm are defined on the Fourier magnitude spectrum, F, as follows:

$$\mathbf{M}_{[n,n]} = \frac{1}{|F(\upsilon,\omega)|}\iint \mathbf{X}_{[n,n]}(\upsilon,\omega)\ F(\upsilon,\omega)\ d\upsilon\, d\omega \qquad (2.1)$$

where $|F(\upsilon,\omega)| = \iint F(\upsilon,\omega)\ d\upsilon\, d\omega$

$$\mathbf{X}_{[1,1]}(\upsilon,\omega) = \begin{pmatrix} \upsilon^2 & \upsilon\omega \\ \upsilon\omega & \omega^2 \end{pmatrix} \qquad (2.2)$$

$$\mathbf{X}_{[2,2]}(\upsilon,\omega) = \begin{pmatrix} \upsilon^4/2 & \upsilon^3\omega/\sqrt{2} & \upsilon^2\omega^2/2 \\ \upsilon^3\omega/\sqrt{2} & \upsilon^2\omega^2 & \upsilon\omega^3/\sqrt{2} \\ \upsilon^2\omega^2/2 & \upsilon\omega^3/\sqrt{2} & \omega^4/2 \end{pmatrix} \qquad (2.3)$$

The eigenvalues of the matrices $\mathbf{M}_{[n,n]}$ are Euclidean invariants [7], which will allow straightforward computation of rotation invariant texture features. Within the scope of this paper, however, we will deal purely with scaling. Under a scalar transformation, α, these matrices are related by

$$\mathbf{M}'_{[n,n]} = \alpha^n \mathbf{M}_{[n,n]} \qquad (2.4)$$

Our strategy is to estimate α as follows, where $\mathbf{m}$ is the vector formed from the elements of $\mathbf{M}$:

$$\alpha = \sqrt[n]{\frac{\|\mathbf{m}_{[n,n]}\|}{\|\mathbf{m}'_{[n,n]}\|}} \qquad (2.5)$$

The moment matrices will be adjusted by this value, α, giving features that are invariant to the particular scaling transformation.

3. ALGORITHM

In the current algorithmic implementation, the mean is removed from each texture sample at the outset, followed by multiplication with a Gaussian window. The former operation provides invariance to grey-scale shifts, while the latter minimises distortion in the spectrum due to the finite image size. Treating the resulting frequency spectra as sampled versions of continuous functions, the integrals of equation 2.1 are thus calculated as discrete summations.

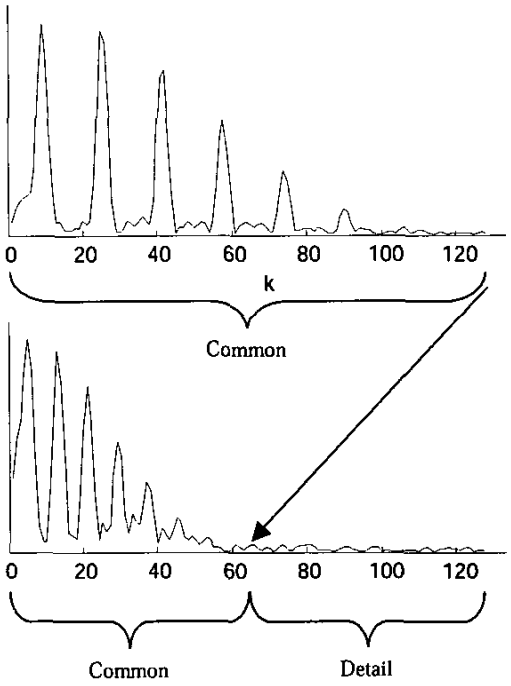

Figure 1. Addition of detail: the upper FFT is zoomed out relative to the lower by a factor of two.

A particular problem arises with the frequency spectra resulting from real zooming operations. As illustrated in figure 1, compression of the spectrum for the "close-up" texture results in new high-frequency components. This corresponds to the introduction of added detail in the zoomed image, not present in the larger scale original. In comparing the spectra of the two texture images, the simple alpha transformation that we have assumed applies only to that segment of the spectrum present in both cases. We need to discount the influence of the detail, since it acts as a confounding factor, by changing the region of integration of equation 2.1 in the case of the zoomed image. This is critical in the context of the moment matrices, which are more sensitive to noise at higher frequencies (increasingly so for higher order moments).

To this end, we propose to truncate the "lower frequency" spectrum based on an estimate of the scaling factor. We assume that the new components at higher frequencies are small in magnitude compared to the "common" section of the spectrum. If this were untrue, it

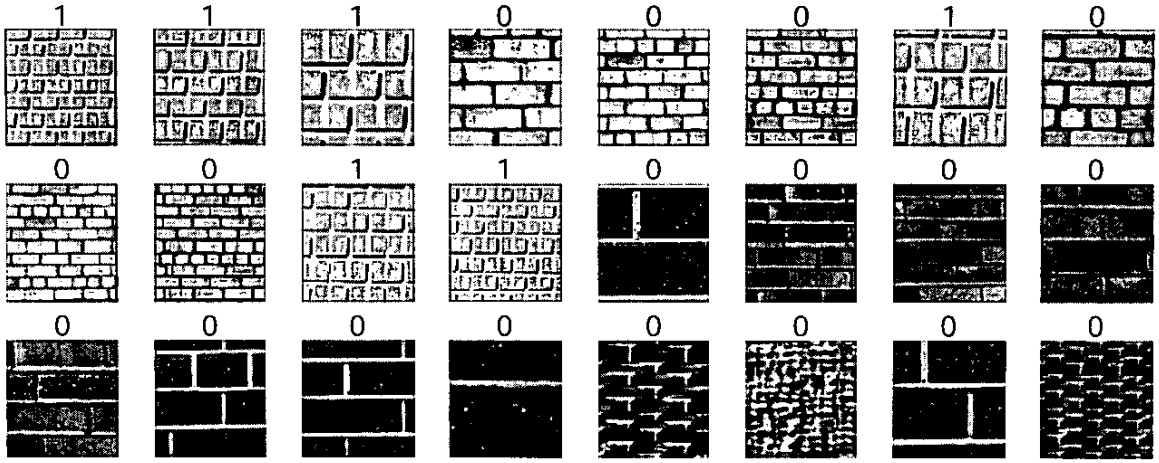

Figure 2. Example retrieval result: relevant results labelled "1".

would mean that a zoomed version of the image would have a significantly different frequency distribution – and would therefore likely be perceived as a dissimilar texture by the human visual system. Thus even when the assumption breaks down, it still captures the spirit of human texture similarity perception. To determine which of the textures is the "lower frequency", or "zoomed in" image, we examine those frequency components above a given magnitude-threshold, as follows:

$$\bar{f} = \frac{\sum_{v,\omega \mid F(v,\omega)>T} F(v,\omega)\sqrt{v^2+\omega^2}}{\sum_{v,\omega \mid F(v,\omega)>T} F(v,\omega)} \qquad (3.1)$$

Where the threshold, T, is specified as a percentage of the peak spectral magnitude. The texture with the smaller value of $\bar{f}$ is identified as the lower-frequency image and the truncation ratio is set as:

$$\text{ratio} = \bar{f}_{low} \Big/ \bar{f}_{high} \qquad (3.2)$$

Equation (3.2) can be interpreted as the ratio of the average radial frequencies. Having estimated this ratio, the lower-frequency distribution is truncated and the moment matrices, $\mathbf{M}_{[1,1]}$ and $\mathbf{M}_{[2,2]}$, calculated for each texture. By truncating the spectrum of the zoomed-in texture image, all frequency components common to both textures are retained. In the current implementation, truncation is radial – i.e. we retain a disk in the frequency plane in each case. Additionally, moments are calculated separately for the two unique quadrants of the FFT, to increase discrimination.

In order to generate the invariant features, the moments of the lower-frequency texture are corrected according to equations (2.4) and (2.5). The Euclidean distance between the moment vectors, $\mathbf{m}_{[2,2]}$, is the distance measure used in our experiment.

Note that, while additional frequency components exist in the spectrum of the zoomed image, the converse effect (from the uncertainty principle and symmetry of the Fourier Transform) is that the zoomed-out image contains extra components spatially. This is perhaps the more obvious effect, since zooming in, by definition, eliminates the periphery. In any case, the effect on the spectrum of the zoomed-out image is to sharpen the peaks (i.e. the frequency resolution is increased). This occurs in figure 1 where, contrary to equation 1.1, the peaks of the upper FFT are not twice as wide as those of the lower.

For equation 1.1 to be satisfied precisely, we should truncate the zoomed-out image in the spatial domain, just as we have truncated the zoomed image in the frequency domain. In practice however, there is little to be gained from this enhancement to the algorithm, as the moment-based approach is inherently robust to some "peak-spreading" in the FFT.

4. EXPERIMENTAL RESULTS

We have tested the distance measure on a set of 112 textures from the Brodatz catalogue [11]. For each of the textures, two non-overlapping regions were used to generate patches at three different scales, giving a total of 6 examples of each texture, or 672 test images in all. The scaled patches were all generated by "zooming out" from a region of the original texture, so as to correctly mimic the effects of zooming in the real world. Starting with a square region of size $1/\alpha$, we resize by a factor of α,

where $\alpha \leq 1$. By allowing only these decimation operations, rather than interpolation (zooming-in, or $\alpha > 1$) we ensure that the "detail" frequency components are indeed present in the zoomed images. The need for this restriction is clear from figure 1: the upper FFT may be deduced from the lower (assuming the texture is spatially homogenous), but not vice versa. Artificial zoom-in cannot generate the detail information.

Tests were conducted using the first 100 of the 672 images as queries to evaluate the accuracy of the scale invariant distance. For each query image, the other 671 images were ranked according to the distance. The average precision vs. recall characteristic, over 100 queries, is given in table 1.

$$\text{recall} = \frac{\#\,\text{relevant matches}\,(k)}{\text{total relevant images}} \tag{3.3}$$

$$\text{precision} = \frac{\#\,\text{relevant matches}\,(k)}{k} \tag{3.4}$$

where k is the rank index.

This is a very demanding test of the algorithm's performance. Although many of the 112 textures are perceptually similar, a "relevant" image, in terms of the precision statistics, is defined as one of the 5 other samples of the query texture. Therefore, no credit is given for finding what may be, to the human eye, very closely matching textures. Since such "credible matches" are a subjective matter, it is difficult to account for them statistically. Figure 2 shows a representative test result. The query image is in the upper left, with the retrieved samples ranked left to right, top to bottom. It can be seen that a number of credibly similar textures have been ranked above relevant patches, degrading the final precision value (to $5/11 \approx 0.45$). In light of this, perhaps table 1 is an unfair representation. Nonetheless, the final precision value of 0.28 indicates that, on average, all five target-samples are found within the top 18 ranks – that is, within the top 3% of the full set.

For comparison, the experiment was repeated using the exact value of the scaling as the truncation ratio in each case. The corresponding results, also given in table 1, demonstrate the potential to improve accuracy if better truncation estimates can be generated.

5. CONCLUSION

We have defined a scale-invariant distance measure for texture recognition applications. This is achieved without the need for log-polar resampling and is based on a general theory of moment invariants. Thus it will lend itself to the calculation of other invariants (e.g. rotation). To increase discrimination, it is possible to include higher-order moments in the calculation. Further work will also include the extension of the techniques described here to texture classification as well as the investigation of more reliable methods to estimate the truncation ratio of equation (3.2) or to eliminate dependency on this parameter altogether. In its current form, the distance measure is also invariant to spatial shifts and grey-scale transform of the texture samples.

Recall	0.2	0.4	0.6	0.8	1.0
Precision	0.78	0.63	0.44	0.36	0.28
Prec. (Perfect Truncation)	0.82	0.72	0.56	0.48	0.41

Table 1. *Precision vs. recall statistics averaged over 100 tests.*

ACKNOWLEDGEMENTS

This work was supported by EPSRC grant GR/M84183, under the Link project Autoarch.

REFERENCES

[1] P.R. Schrater, D.C. Knill and E.P. Simoncelli, "Perceiving Visual Expansion Without Optic Flow", Nature, Vol. 410, pp 816-819, 12 April 2001.

[2] S. Derrode and F. Ghorbel, "Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description", Computer Vision and Image Understanding, Vol. 83, No. 1, July 2001.

[3] H. Xiong, T. Zhang and Y.S. Moon, "A Translation- and Scale-Invariant Adaptive Wavelet Transform", IEEE Trans. Image Processing, Vol. 9, No. 12, Dec. 2000.

[4] F.S. Cohen, Z.G. Fan and M.A. Patel, "Classification of Rotated and Scaled Textured Images using Gaussian Markov Random Field Models", IEEE Trans. PAMI, Vol. 13, No. 2, pp 192-202, Feb. 1991.

[5] H. Potlapalli and R.C. Luo, "Fractal-Based Classification of Natural Textures", IEEE Trans. Industrial Electronics, Vol. 45, No. 1, pp 142-150, Feb. 1998.

[6] B. Julesz, "Experiments in the Visual Perception Of Texture", Scientific American, 232, pp 34-43, 1975.

[7] G. Taubin and D.B. Cooper, "Object Recognition Based on Moment (or Algebraic) Invariants", in J. Mundy and A. Zisserman, eds., *Geometric Invariance in Computer Vision*, MIT Press, Cambridge, Mass., pp 375-397, 1992.

[8] R.J. O'Callaghan and D.R. Bull, "Improved illumination-invariant descriptors for robust colour object recognition", accepted for publication, ICASSP 2002.

[9] G. Healey and D. Slater, "Global color constancy: recognition of objects by use of illumination invariant properties of color distributions", J. Optical Society of America A, Vol. 11, No. 11, pp 3003-3010, Nov. 1994.

[10] Y. Yoshida and Y. Wu, "Classification for Rotated and Scaled Textured Images Using Invariants based on Spectral Moments", IEICE Trans. Fundamentals, Vol. E81-A, No. 8, August 1998.

[11] P. Brodatz, "Textures, A Photographic Album for Artists and Designers", Dover Publications, New York, 1966.